Check for updates

Journal of Chemometrics

WILEY

Journal of CHEMOMETRICS

RESEARCH ARTICLE OPEN ACCESS

Handling Experimental Data in FACPACK: Noise Effects on Duality Relations and Algorithmic Parameter Selection

Mathias Sawall[1] | Christoph Kubis[2] | Alejandro Olivieri[3] | Steven J. Roeters[4] | Klaus Neymeyr[1,2]

[1]Institut für Mathematik, Universität Rostock, Rostock, Germany | [2]Leibniz-Institut für Katalyse, Rostock, Germany | [3]Instituto de Química Rosario (CONICET-UNR), Facultad de Ciencias Bioquímicas y Farmacéuticas, Universidad Nacional de Rosario, Rosario, Argentina | [4]Department of Anatomy and Neurosciences, Amsterdam UMC, Amsterdam, the Netherlands

Correspondence: Mathias Sawall (mathias.sawall@uni-rostock.de)

Received: 7 April 2026 | **Revised:** 28 August 2026 | **Accepted:** 1 September 2026

Keywords: area of feasible solutions | duality | experimental and noisy data | FACPACK | feasible bands

ABSTRACT

The area of feasible solutions (AFS) quantifies factor ambiguity in multivariate curve resolution (MCR). Although AFS computation is well-defined for noise-free data, experimental datasets require constraint relaxation and parameter tuning, which influence the resulting feasible bands. This work addresses the computation of the AFS in the presence of noise. We propose a statistically motivated strategy for control parameter estimation based on Savitzky–Golay filtering, introduce a two-stage preprocessing approach for baseline-distorted data, and analyze the effect of noise in the dual space. The methods are demonstrated on model data and experimental FT-IR datasets.

1 | Introduction

For bilinear spectroscopic data collected in the matrix D, the multivariate curve resolution (MCR) problem consists in determining a concentration matrix C and a spectral matrix S such that

$$D = CS^{\mathrm{T}} + E,$$

where E represents residual errors [1, 2]. Here, $D \in \mathbb{R}^{k\times n}$ contains k mixed spectra measured at n frequency channels, $C \in \mathbb{R}^{k\times s}$ contains the concentration profiles, and $S \in \mathbb{R}^{n\times s}$ contains the pure component spectra, with $s \ll \min(k, n)$ denoting the number of chemical species.

A major difficulty of this factorization is its non-uniqueness even after excluding trivial scaling and permutation ambiguities. Often a continuum of feasible profile pairs (C, S) exists [3, 4]. Characterizing this continuum is essential for assessing the reliability of a particular MCR solution, but its computation is nontrivial.

Several approaches have been developed to quantify rotational ambiguity. Profile-based methods include MCR-BANDS [5–7] and sensorwise NBANDS [8, 9]. Geometric approaches comprise Borgen–Rajkó plots [10, 11] and generalized Borgen plots [12]. Numerical approximation-oriented approaches for an area of feasible solutions (AFS) computation are triangle enclosure [13, 14], polygon inflation [15, 16], ray casting [17] and sampling methods [18, 19]. The AFS provides a low-dimensional representation of all admissible profiles and thus quantifies the extent of rotational ambiguity.

For noise-free systems with a small number of components, the AFS can be computed efficiently [4, 10, 20]. In the presence of noise, however, the strict nonnegativity and/or exact factorization constraints must be relaxed. This relaxation is implemented via control parameters and thresholds that determine the admissibility of the solutions. MCR-bands, sensorwise NBANDS, and triangle enclosure relax the factorization constraint, whereas the FACPACK implementations polygon inflation and ray casting relax the nonnegativity constraint.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2026 The Author(s). *Journal of Chemometrics* published by John Wiley & Sons Ltd.

The FACPACK software [21] includes three packages, which can be used to approximate the AFS for noise-free model data, as well as for experimental, noisy data: polygon inflation, generalized Borgen plots, and ray casting. Polygon inflation and ray casting relax the nonnegativity constraints and optimize a cost function to minimize the amount of negative entries, as described in [22–24]. Control parameters are applied to decide whether a noisy data factorization is acceptable. For the approximation of an AFS, an appropriate choice of control parameters and a good signal-to-noise ratio in the data are much more important than for the calculation of a 'single' MCR factorization without a simultaneous factor ambiguity analysis.

1.1 | Purpose of this Work

This work investigates the computation and application of the AFS and duality for noisy data. The focus is on the algorithmic and numerical aspects affecting the reliability of feasible sets and derived bands.

1. How can control parameters be selected in a systematic and reproducible manner? If they are chosen too small, admissible and chemically meaningful profiles may be excluded. If chosen too large, the constraints may lose discriminative strength.
2. How can residual baseline distortions be corrected when an initial baseline treatment still prevents a reliable approximation of the AFS?
3. How can bands of confidence be derived by using the duality principles when a pure component spectrum is affected by noise? Duality imposes restrictions on one factor based on additional information about the other. A case study analyzes the impact of noise and identifies conditions under which dual constraints are useful for reducing of factor ambiguity.

1.2 | Assumptions on the Type of Noise and Limitations

This paper presents procedures and guidelines for handling experimental and noisy data in the context of an ambiguity analysis. The focus is on the strategy relaxing the nonnegativity constraint, as applied in FACPACK. If the noise distribution is known, statistical arguments can be used to determine suitable control parameters. Our analysis assumes random noise and does not account for systematic distortions. Therefore, the proposed parameter choices should be regarded as recommendations. Experimental data may contain both random and systematic perturbations and may require case-specific adjustments. The proposed procedures are intended to support the parameter selection and subsequent data treatment. They are not designed as automatic routines that yield reliable results without user inspection. The experimental datasets illustrate typical situations. In one example, the statistically suggested thresholds are too restrictive and must be increased, demonstrating that manual adjustment can be necessary.

1.3 | Structure of the Paper

The paper is organized as follows. Section 2 introduces the penalty functions and control parameters used in FACPACK for MCR computations and outlines the approximation of the AFS for experimental data. Section 3 addresses the first question posed in Section 1.1, namely, the estimation of control parameters and thresholds based on statistical considerations. Section 4 addresses the second question and presents a two-stage procedure for improving insufficient baseline corrections to enable reliable AFS computations. Section 5 analyzes the effect of noise in duality-based investigations (Question 3 in Section 1.1). Numerical results are presented directly after the corresponding theoretical developments in each section.

1.4 | Datasets

The following three datasets (two experimental and one model) are used to illustrate the proposed ideas.

Dataset 1. The FTIR-spectroscopic dataset is about tracking the formation and activation of inactive polynuclear rhodium carbonylcomplexes from $Rh(acac)(CO)_2$ as a precatalyst for the hydroformylation reaction in the absence of phosphorus ligands [25]. Under typical hydroformylation conditions at 100°C, 20 bar of synthesis gas ($CO/H_2 = 1{:}1$), with dodecane as a solvent, the starting material $Rh(acac)(CO)_2$ with an initial concentration of $1 \cdot 10^{-3}$ mol/L is progressively transformed to $Rh_4(CO)_{12}$ and $Rh_6(CO)_{16}$ in a consecutive manner. The data contains absorption values for $k = 208$ spectra and $n = 1453$ wavenumbers in the range $[1800,\ 2150]\ \mathrm{cm}^{-1}$. Absorption signals originate from the $s = 3$ named species. The dataset has a high signal-to-noise ratio and sharp, partly overlapping peaks. It is suitable for studying the impact of noise. Figure 1 shows the mixed profiles and the corresponding pure component profiles. This dataset has many small negative entries due to background subtraction, cf. Section 2.3.

Dataset 2. The infrared spectral data stem from a research on the hydroformylation of neohexene with a catalyst system based on $Rh(acac)(CO)_2$ as rhodium precursor and tris(2,4-di-tert-butylphenyl)phosphite (TDTBPP) as bulky monophosphite ligand [26, 27]. The FTIR spectra were collected during the monitoring of the reaction at the following conditions: q = 30° C, P(CO/H2) = 20 bar, [neohexene] = $500 \cdot 10^{-3}$ mol/L, [Rh] = $0.3 \cdot 10^{-3}$ mol/L, [TDTBPP] = $6 \cdot 10^{-3}$ mol/L, solvent: n-hexane. The dataset contains $k = 1050$ spectra with $n = 665$ wavenumbers in the interval $[1960, 2120]\ \mathrm{cm}^{-1}$, and the $s = 3$ absorbing species olefin, acyl complex and hydrido complex. The mixed spectra and the computed pure profiles are shown in Figure 2.

Dataset 3. The model dataset is based on the model kinetic $X \to Y \to Z$ with the reaction rate constants $\kappa_1 = 1$, $\kappa_2 = 0.5$, and the initial concentration vector $C(1,:) = (1, 0, 0)$. The time span is $[0, 7]$, and $k = 300$ nodes are used. The spectra are modeled using sums of Gaussians as

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$s_X(\lambda) = 0.3\exp\left(-\frac{(\lambda-115)^2}{100}\right)+\exp\left(-\frac{(\lambda-180)^2}{130}\right),$$
$$s_Y(\lambda) = \exp\left(-\frac{(\lambda-120)^2}{50}\right)+0.75\exp\left(-\frac{(\lambda-178)^2}{100}\right),$$
$$s_Z(\lambda) = 0.5\exp\left(-\frac{(\lambda-105)^2}{100}\right)+\exp\left(-\frac{(\lambda-175)^2}{130}\right)$$

with $n = 500$ equidistant nodes for $\lambda \in [100, 200]$. The presented ideas are evaluated using the model data with and without noise. Homoscedastic noise is simulated by adding normally distributed random numbers with mean 0 and standard deviation 0.015. Figure 3 shows the noisy data and the original pure component profiles.

2 | MCR Solutions for Experimental and Noisy Data

In general, perturbations complicate both the analysis of factor ambiguity and, in some cases, even the determination of a single factorization. For noisy data, the exact factorization condition must be replaced by a best approximation, so that $E \neq 0$. However, it is often impossible to obtain factors that are simultaneously nonnegative and minimize the residual norm $\|E\|_F$. Consequently, one of the two requirements, *best approximation* or *nonnegativity*, must be relaxed.

Methods such as MCR-ALS [28], MCR-BANDS [5–7], sensor-wise NBANDS [8], and classical NMF algorithms [29, 30] relax

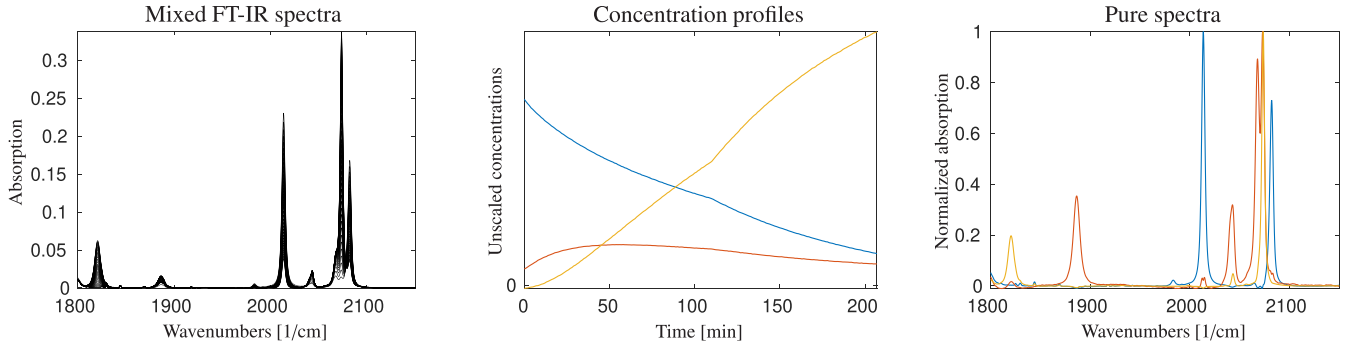


FIGURE 1 | The FT-IR data (left) from dataset 1 with a high signal-to-noise ratio and the pure component profiles underlying C (center) and S (right).

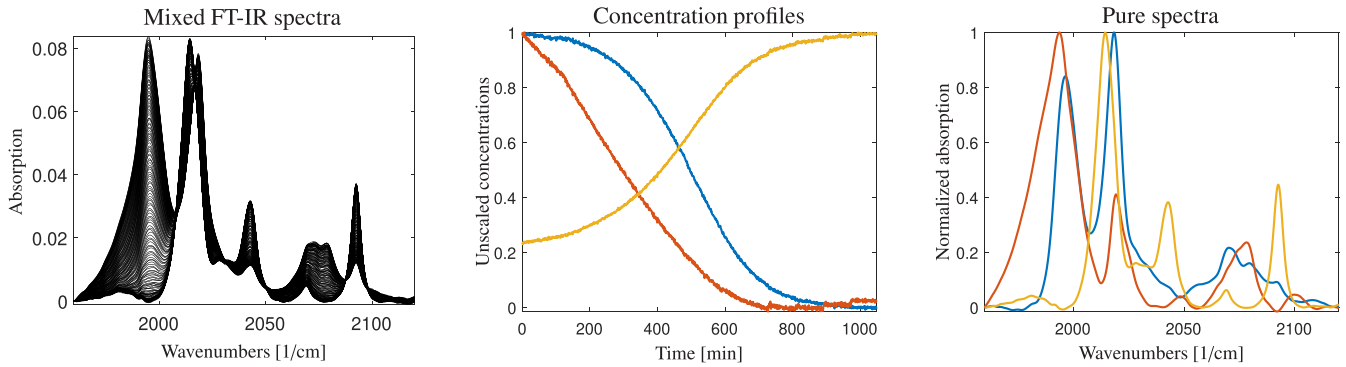


FIGURE 2 | The FT-IR data (left) after a two-stage baseline correction and the pure component concentration profiles C (center) and pure spectra S (right) for dataset 2.

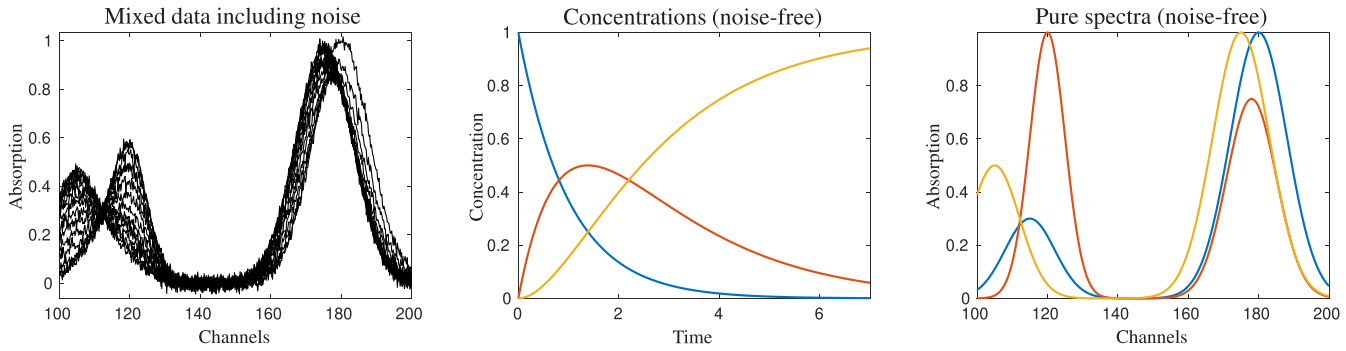


FIGURE 3 | The data of the model problem from dataset 3. Left: the mixed spectra including noise (only every 20th plotted). Center and right: The concentration profiles and the pure spectra without noise.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the best-approximation condition. In contrast, BTEM [24] and the FACPACK routines polygon inflation [15, 31] and ray casting [17] relax the nonnegativity constraints. See [1, 2] for general aspects of MCR for noisy data, and [32] for differences when applying soft or hard constraints in MCR computations.

Because this paper addresses control parameter selection in FACPACK, the relaxation of nonnegativity constraints is considered in more detail. The factorization is based on a singular value decomposition (SVD), which extracts the main structure-determining information from the mixed data and provides two bases for reconstructing C and $S^{\mathbf{T}}$. The truncated SVD guarantees that $CS^{\mathbf{T}}$ is the best rank-s approximation of D, i.e.,

$$\min_{C\in\mathbb{R}^{k\times s},\ S\in\mathbb{R}^{n\times s}} \underbrace{\left\|D-CS^{\mathbf{T}}\right\|_F^2}_{=ssq(D-CS^{\mathbf{T}})} = \sum_{\ell=s+1}^{\min(k,n)} \sigma_\ell^2$$

with σ_ℓ being the ℓth singular value of D, see [33, 34].

2.1 | SVD-Based MCR-Computations

Let U, Σ, and $V^{\mathbf{T}}$ denote the factors of the s-dimensional truncated SVD of D, i.e., U and V contain the first s left and right singular vectors, and $\Sigma\in\mathbb{R}^{s\times s}$ contains the corresponding singular values. Assuming $CS^{\mathbf{T}}=U\Sigma V^{\mathbf{T}}$, the factorization problem reduces to determining a regular matrix $T\in\mathbb{R}^{s\times s}$ such that

$$C=U\Sigma^{1/2}T^{-1},\qquad S^{\mathbf{T}}=T\Sigma^{1/2}V^{\mathbf{T}}, \tag{1}$$

see [1, 2]. Here, the singular values are distributed symmetrically to both factors via $\Sigma^{1/2}$. This scaling avoids distortions of the coordinate axes in the AFS representation.

If $\mathrm{rank}(T)=s$, the SVD construction guarantees that $CS^{\mathbf{T}}$ remains a best rank-s approximation of D, i.e.,

$$\left\|D-CS^{\mathbf{T}}\right\|_F\to\min.$$

In numerical factorization, nonnegativity of C and S can be enforced, or negative entries penalized, by suitable penalty functions. A single MCR solution (c,S) can be calculated by minimizing

$$f(T)=\|\min(C,0)\|_F^2+\|\min(S,0)\|_F^2+\|I-T^{+}T\|_F^2, \tag{2}$$

where the first two terms penalize negative entries in C and S, and the last term rules out singular matrices T. Therein, T^{+} is the pseudoinverse of T.

2.2 | The AFS for Noise-Free Data

The AFS can be understood as a 'global' approach to the factor ambiguity problem because it represents all feasible nonnegative profiles. For an s-component system, the AFS is a bounded subset of $\mathbb{R}^{s-1}$ and constitutes a low-dimensional representation of all profiles occurring in factorizations of a given matrix D. For noise-free model data, the AFS in V-space is defined by

$$\mathscr{M}=\left\{x\in\mathbb{R}^{s-1}:\ \text{exists a regular } T\in\mathbb{R}^{s\times s} \text{ with } T(1,:)=(1,x^{\mathbf{T}}) \text{ and } C,S\geq 0\right\}. \tag{3}$$

Details on the AFS and its computation are given in [4, 20].

The geometric construction of the AFS relies on an outer and an inner polytope [10–12]. In the V-space, with the singular values distributed symmetrically via $\Sigma^{1/2}$ to the left and right singular vectors, the outer polytope $\mathscr{F}_S$ and the inner polytope $\mathscr{I}_S$ are given by

$$\mathscr{F}_S=\left\{x\in\mathbb{R}^{s-1}:\ (1,x^{\mathbf{T}})\Sigma^{1/2}V^{\mathbf{T}}\geq 0\right\},$$
$$\mathscr{I}_S=\mathrm{convhull}(\{a_1,\ldots,a_k\}),$$

where the data points $a_i\in\mathbb{R}^{s-1}$ are defined as

$$a_i=\frac{(U\Sigma^{1/2})^{\mathbf{T}}(2{:}s,i)}{(U\Sigma^{1/2})^{\mathbf{T}}(1,i)},\quad i=1,\ldots,k. \tag{4}$$

The two polytopes encode the nonnegativity constraints of the factors and are coupled by duality to the corresponding polytopes $\mathscr{F}_C$ and $\mathscr{I}_C$ in the U-space [16, 35–37]. In the V-space, the outer polytope corresponds to the nonnegativity of S, whereas the inner polytope represents the nonnegativity of C. If we use the polytope notation without the factor-related index, then it refers to the V-space.

2.3 | The AFS for Noisy Data

Although the objective function in (2) is suitable for computing MCR factorizations, its direct use is not appropriate for deciding feasibility within the AFS in the presence of noise. For noisy data, small negative entries must be tolerated. We therefore introduce relative thresholds ϵ_C and ϵ_S that define admissible deviations from nonnegativity. Negative entries are accepted if their relative magnitude does not exceed the prescribed threshold. This strategy is implemented in polygon inflation and ray casting [15, 17]. The relaxed conditions read

$$\frac{C(i,j)}{\max(C(:,j))}\geq-\epsilon_C,\quad \frac{S(\ell,j)}{\max(S(:,j))}\geq-\epsilon_S \tag{5}$$

with $\epsilon_C,\epsilon_S\geq 0$. A factorization is classified as *feasible* if and only if (5) holds for all $i=1,\ldots,k$, $j=1,\ldots,s$, and $\ell=1,\ldots,n$. Thus, each acceptable negative entry must remain within the prescribed relative bound.

Noise and negative entries directly affect the inner and outer polytopes. Under strict constraints $C,S\geq 0$, the AFS may become empty. In datasets with sharp peaks and spectral regions of negligible absorption, it may occur that the inner polytope is not contained in the outer polytope because some data points lie outside $\mathscr{F}$; this always happens when D has negative entries. Figure 4 illustrates this for the FT-IR data from dataset 1. Although the data exhibit a high signal-to-noise ratio and no clearly visible baseline distortions, many data points lie outside the outer polytope in U-space. In V-space, the outer polytope is empty due to inconsistencies between the nonnegativity

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

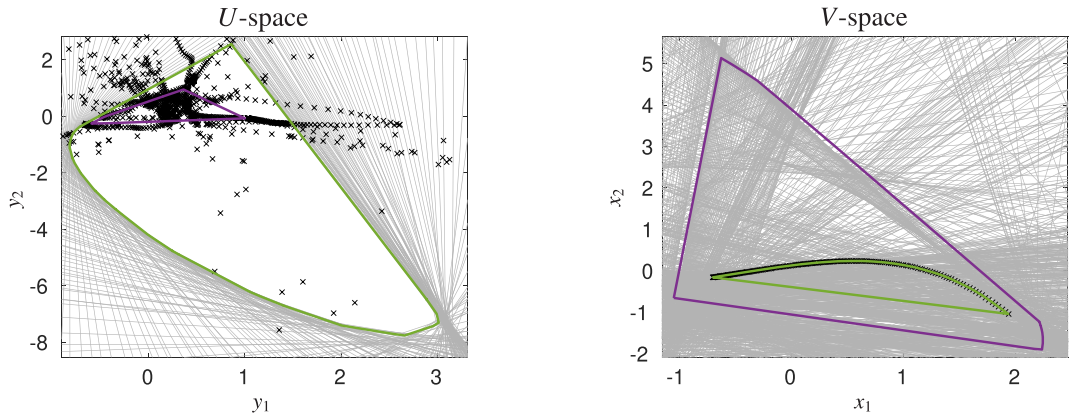


FIGURE 4 | The major difficulty in computing the AFS for dataset 1 is caused by noise. Although the FT-IR data exhibit a high signal-to-noise ratio and the pure profiles contain only small negative values, the AFS becomes empty under strict nonnegativity constraints. The gray lines indicate the exact nonnegativity restrictions, and the black crosses denote the data points. No factorization with nonnegative factors exists under these conditions. In the U-space, many data points lie outside $\mathscr{F}_C$, and in V-space the outer polytope $\mathscr{F}_S$ is an empty set. When the nonnegativity constraints are relaxed slightly, the modified outer and inner polytopes are obtained. They are shown in green ($\mathscr{F}_C$ and $\mathscr{I}_S$) and purple ($\mathscr{F}_S$ and $\mathscr{I}_C$), respectively. The control parameters are $\varepsilon_C = 1.15 \cdot 10^{-2}$ and $\varepsilon_S = 1.04 \cdot 10^{-3}$, see Section 3.1.

constraints and the reconstruction condition to work with only $s = 3$ right singular vectors.

To obtain meaningful approximations of $\mathscr{F}$ and $\mathscr{I}$, the constraints for the polytope constructions must be relaxed. Condition (5) slightly enlarges the outer polytope compared with the strict constraint $S \geq 0$, whereas the inner polytope is reduced accordingly.

2.4 | Importance of Suitable Control Parameters

In polygon inflation and ray casting, the feasibility of a given $x \in \mathbb{R}^{s-1}$ is tested by minimizing a cost function. This function contains penalty terms enforcing the relaxed nonnegativity conditions in (5) as well as the regularity of T. The same principle applies to noise-free and noisy data.

For an efficient implementation, the cost function is constructed such that only points with a value close to zero are accepted as feasible. This allows a rapid preliminary test of the necessary condition $x \in \mathscr{F}$, see [31]. In FACPACK, a threshold of 10^{-10} is used to classify a point x as feasible [15].

The parameters ε_S and ε_C are the only quantities controlling the admissible noise level. Their choice is therefore critical. If they are too small, no solution is obtained because the SVD-based factorization does not satisfy (5) exactly. If they are too large, the constraints lose their restrictive effect and the AFS no longer provides meaningful information.

2.5 | Alternative Relaxations and Hard Constraints

As for the computation of an individual MCR-solution, both, soft and hard constraints can be used to approximate the AFS [32, 38]. Both approaches have advantages and disadvantages. The application of nonnegativity constraints in (5) in the AFS computation combines soft and hard constraints. On the one hand, the restrictions are implemented using soft constraints. A least squares minimization via penalty terms is employed and relative entries smaller than $-\varepsilon_C$, respectively $-\varepsilon_C$, are not set to zero but generate high values in the penalty functions. On the other hand, however, the AFS calculation only accepts points that generate a least squares error of at most 10^{-12}. Thus, it is more or less a hard constraint. In particular, the control parameters ε_C and ε_C for permissible relative negative entries are of enormous importance, as they effectively set the limit for active constraints. Furthermore, this approach ensures the important point that the lack-of-fit is constant and minimal.

When approximating the AFS for noisy data, negative entries in the computed factors are the main problem. In contrast to applying (5), another common strategy is to truncate negative entries by setting them to zero and relax the best-approximation constraint instead. Defining $C_+ = \max(C, 0)$ and $S_+ = \max(S, 0)$, the residual $D - C_+ S_+^{\mathbf{T}}$ is then used to assess the factorization, typically via a sum-of-squares criterion. This approach is applied in [5–8, 13]. As in the present paper, the choice of a suitable threshold for $\| D - C_+ S_+^{\mathbf{T}} \|$ remains critical, see [9, 39]. However, this strategy does not permit the rapid feasibility test available in polygon inflation and ray casting. The rapid test is one of the main reasons why the numerical approximation of the AFS for experimental data can be done much more quickly than with hard constraints.

Another possible modification is to truncate negative entries in D itself. This aims to prevent negative values in C and $S^{\mathbf{T}}$ from the beginning. Such preprocessing is problematic. For $s \geq 3$, it often does not solve the problem of the absence of a nonnegative factorization. Moreover, profiles may be affected by positive noise contributions. In spectral regions where mixed signals are well above zero, negative contributions of one component cannot be removed independently because they are masked by others.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3 | Determination of Proper Control Parameters

This section presents a procedure to determine suitable control parameters for the relaxed nonnegativity condition in (5). The approach is based on statistical considerations and is illustrated later by two examples.

We assume homoscedastic noise that is independent and normally distributed with mean 0 and standard deviation σ, which is critical for some applications but proper for IR measurements. Under this assumption, sample variances and quantiles provide a basis for selecting appropriate values of ε in (5). For a normally distributed random variable X with mean 0 and standard deviation σ, the probability that $X < -\sigma$ is approximately 15.9 %, see [40]. Only about 0.6 % of realizations satisfy $X < -2.5\sigma$, see Figure 5. Hence, if $\varepsilon_S = -2.5\sigma$, only a small fraction of noise-induced violations leads to a truncation of the outer polytope. The remaining question is how to estimate the standard deviation σ of a noise level from experimental data.

3.1 | Statistical Approximation of the Standard Deviation

To estimate ε_S, we apply a data smoothing procedure in frequency direction. The parameter ε_C is obtained analogously from $D^{\mathbf{T}}$. Because the AFS is computed from the first s singular vectors, the truncated SVD already reduces noise. Nevertheless, some negative entries often remain in the rank-s approximation and may prevent a nonnegative factorization. To distinguish clearly between original and filtered data, let $D^{(\mathrm{SVD})}$ denote the rank-s approximation of D.

We estimate the noise level by smoothing each profile in $D^{(\mathrm{SVD})}$ with a Savitzky–Golay (SG) filter [41] and analyze the distances between the non-SG-smoothed data and the smoothed data. Assuming normally distributed noise with mean zero, its standard deviation is approximated by unbiased local sample variances computed over the filter windows. A similar multi-stage smoothing method to handle noisy data is also used in [5].

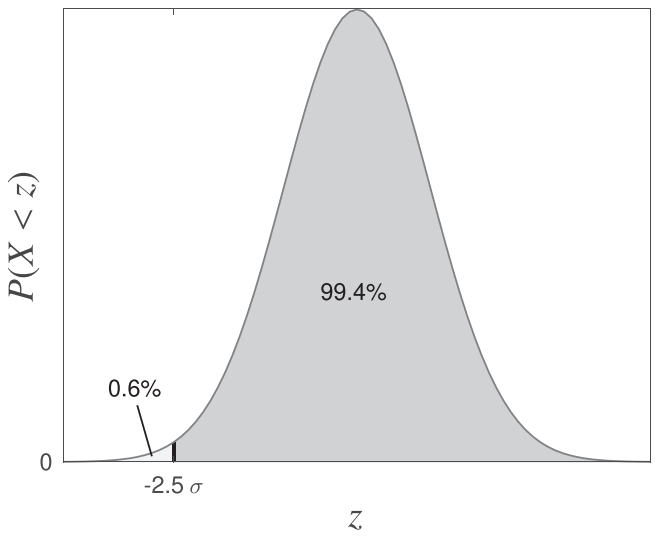


FIGURE 5 | For a normally distributed random variable $X \sim \mathcal{N}(0, \sigma^2)$, the probability of $X < -2.5\sigma$ is approximately $6 \cdot 10^{-3}$.

Let $D^{(\mathrm{SG})}$ denote the data after SG-smoothing in frequency direction. Choose $n_l \in \mathbb{N}$, for example $n_l \in \{5, 10\}$, and fit a cubic polynomial in a least-squares sense to the window $D^{(\mathrm{SVD})}(i, j-n_l{:}j+n_l)$. Evaluating the polynomial at the center yields $D^{(\mathrm{SG})}_{ij}$. Define $E = D^{(\mathrm{SVD})} - D^{(\mathrm{SG})}$. For $i = n_l, \ldots, n-n_l$ and $j = 1, \ldots, k$, the local unbiased sample variances are

$$v_{ij}^2 = \frac{1}{2n_l} \sum_{\ell=j-n_l+1}^{j+n_l} \left(E_{i\ell} - m_{ij}\right)^2 \tag{6}$$

with local means

$$m_{ij} = \frac{1}{2n_l+1} \sum_{\ell=j-n_l+1}^{j+n_l} E_{i\ell}.$$

To match the normalization in (5), we average the local standard deviations and scale them by the maximal data magnitude,

$$\delta_S = \frac{1}{k(n-2n_l)\max_{ij}(D_{ij})} \sum_{i=1}^{k} \sum_{j=n_l}^{n-n_l} v_{ij}. \tag{7}$$

For experimental data, it is often useful to combine this statistical estimate with the largest negative entry in $D^{(\mathrm{SVD})}$. We therefore define

$$\varepsilon = \frac{1}{2}\left(\frac{\left|\min\left(0, \min\left(D^{(\mathrm{SVD})}\right)\right)\right|}{\max\left(D^{(\mathrm{SVD})}\right)} + 2.5\delta_S \right). \tag{8}$$

The factor 2.5 reflects that only about 0.6 % of normally distributed values are expected below -2.5σ. The same procedure is applied to determine ε_C.

Remark 1.

1. Smoothing the data and computing v_{ij} are critical steps. When computing v_{ij} for ε_S we compare the error values $E_{i\ell}$ in (6) with the mean of the error values of the neighboring points in frequency direction. Thus, the comparison is done in the same direction in which the smoothing was applied. For the computation of ε_C the data is smoothed in time direction, and the neighboring values in time direction are taken into account. This works well with fine resolutions and when the absorption behavior is similar in both the time and frequency directions. Swapping the rows and columns for the comparative calculation would also yield meaningful results. If only one control parameter is to be used for both directions as $\varepsilon = \varepsilon_C = \varepsilon_S$, it also makes sense to include all neighboring values on a 2D grid in the mean calculation.
2. The procedure may also be applied without initial smoothing. Smoothing the data is the basic step to bring the method for very different datasets to a reasonably common starting point.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.2 | Application to Dataset 1

The proposed strategy is applied to the experimental data from dataset 1. To estimate the standard deviation, an SG filter with $n_l = 10$ is used in frequency and time direction. Each cubic polynomial is fitted to 21 data points. The resulting estimates for $2.5\delta_{C/S}$ and the corresponding control parameters are $2.5\delta_S = 4.36 \cdot 10^{-3}$, $\varepsilon_S = 2.88 \cdot 10^{-3}$, $2.5\delta_C = 6.88 \cdot 10^{-4}$, $\varepsilon_C = 1.04 \cdot 10^{-3}$. However, the spectral AFS then only includes two true profiles. Because ε_S is smaller than $2.5\delta_S$, the value of ε_S is multiplied by 4. These values yield the relaxed outer and inner polytopes (colored lines) shown in Figure 4. Under strict nonnegativity, the corresponding AFS sets are empty.

3.3 | Application to Dataset 2

The same procedure is applied to dataset 2. Again, strict nonnegativity leads to empty AFS sets. The statistical estimates give $\varepsilon_S = 1.84 \cdot 10^{-2}$ and $\varepsilon_C = 2.52 \cdot 10^{-2}$, but these values still result in empty feasible sets. Larger thresholds are required; in this example, four times the estimated values are used. The resulting polytopes are shown in Figure 6. For questions regarding the AFS sets for various control parameters ε_C and ε_S we refer to [4], which presents such plots for a similar dataset, and [12, 15] in general. Increasing the value of ε_C increase $\mathscr{F}_C$ and decrease $\mathscr{I}_S$. Thus, it typically increases both, the spectral and the concentration AFS. The the impact of the value of ε_S is analogous.

This dataset shows that the statistical procedure provides guidance but not definitive parameter values. Manual adjustment may be necessary. A closer inspection indicates that the data require additional preprocessing to improve the baseline. This topic is addressed in Section 4.

3.4 | Application to the Model Dataset 3

This section deals with the model dataset 3. Our aim is to demonstrate that the proposed estimation of the control parameters yields meaningful results for noisy data. Because the AFS sets and feasible bands of the noise-free data are known, they serve as references.

For the noisy data, (7) results in $\delta_C = 0.0141$ and $\delta_S = 0.0126$. These values are consistent with the imposed normally distributed noise of mean 0 and standard deviation 1.5%. The minimal relative entry in the rank-3 reconstruction is

$$\frac{\min\,(D^{(\mathrm{SVD})})}{\max\,(D^{(\mathrm{SVD})})} = -1.49 \cdot 10^{-2}.$$

Using (8), this leads to $\varepsilon_C = 2.4 \cdot 10^{-2}$ and $\varepsilon_S = 2.5 \cdot 10^{-2}$. With these parameters, the AFS sets and feasible bands are approximated reliably.

Figure 7 compares the AFS sets of the noise-free data ($\varepsilon_C = \varepsilon_S = 10^{-12}$) and the noisy data with the estimated parameters. The sets are similar but not identical. The displacement of the data points illustrates the influence of noise. In this example, the points in U-space are more affected than those in V-space. Figure 8 shows the corresponding feasible bands. The differences between noise-free and noisy data are small. Obviously, the feasible profiles of the original data are included in the set of feasible profiles of the noisy data. Conversely, some feasible profiles of the noisy dataset are not included in the set of feasible profiles of the original data without noise. The main differences are in the absolute values, but the shapes of the profiles are very similar. This difference is to be expected, as slightly negative entries are taken into account. These results indicate that the proposed estimation of ε_C and ε_S provides useful and realistic approximations of the rotational ambiguity.

3.5 | Variation of the Smoothing Parameter

The integer n_l defines the size of the window for smoothing the data as initial step when computing the control parameters. We apply a least-squares approximation of the data including n_l left and n_l right neighboring values. Using a polynomial of degree 3 requires at least $n_l \geq 2$. Next, we vary the value of n_l for the model dataset 3 and compare the resulting values for

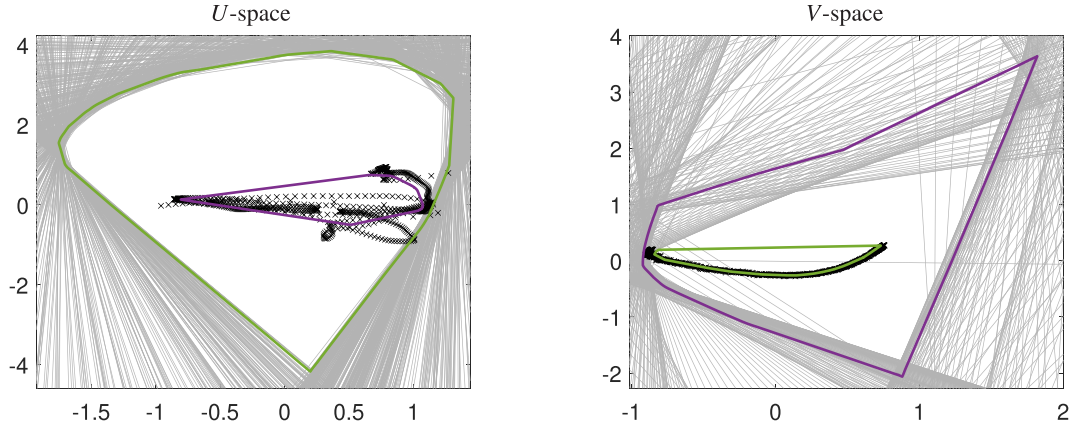


FIGURE 6 | The low-dimensional representation in the U- and V-space for relaxing the nonnegativity constraints are computed for the dataset 2. The resulting values are $\varepsilon_S = 1.84 \cdot 10^{-2}$ and $\varepsilon_C = 2.52 \cdot 10^{-2}$, see Section 3.1. These relaxed nonnegativity conditions enable a reasonable AFS approximation for the spectra and the concentration factors. Without relaxing the constraints, the AFS sets would be empty (gray lines and the black stars). The relaxed inner and outer polytopes in the U- and V-space are plotted in purple and green, respectively, using the same color code as in Figure 4.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

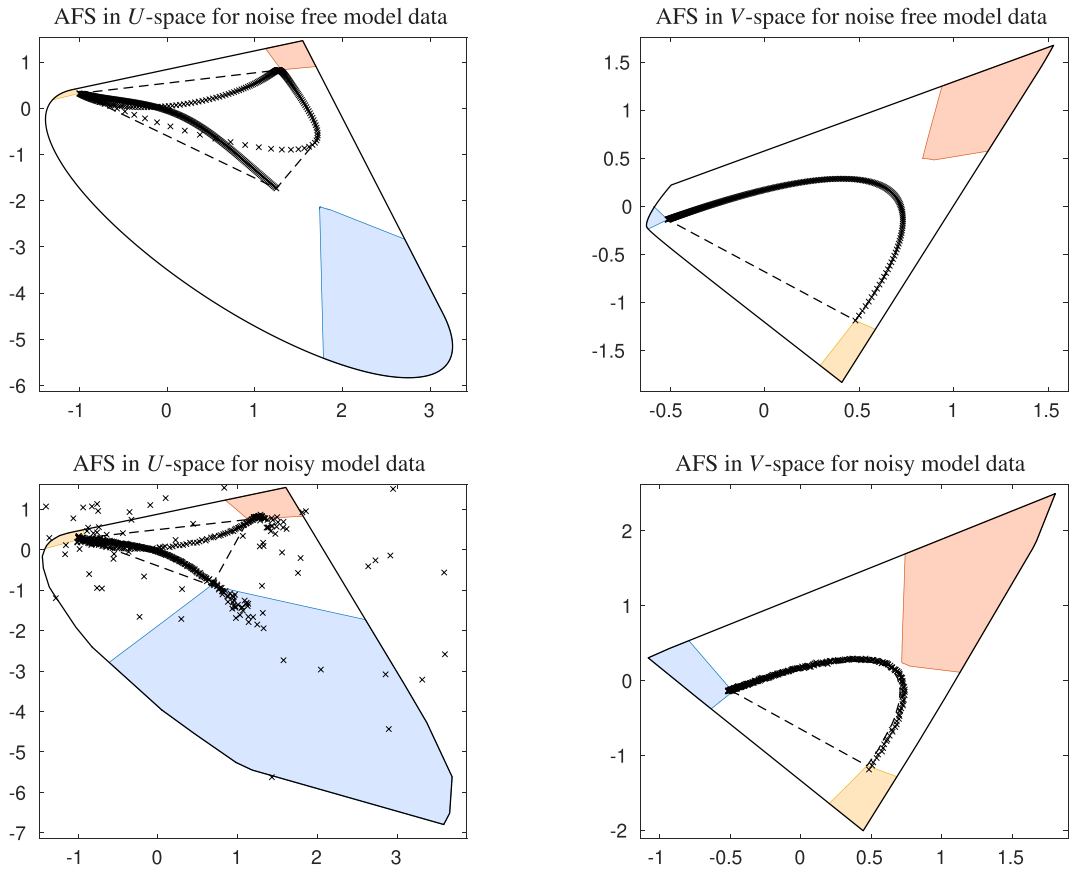


FIGURE 7 | The AFS sets (colored areas) for the model data from dataset 3 are shown for the noise-free data in the top row and for the noisy data in the second row. The control parameters used to approximate the AFS sets for the noisy data are $\varepsilon_C = 2.4 \cdot 10^{-2}$ and $\varepsilon_S = 2.5 \cdot 10^{-2}$. The outer polygons are plotted by black lines and the inner polygons by broken lines. The data points are plotted as black crosses. The results are similar but not equal.

the control parameters, as well as the approximations of the AFS sets, to analyze its impact. The selection of a proper n_l mostly depends on the resolution of the spectrometer compared with the width of the peaks, see also Remark 1. Figure 9 shows the results for $n_l \in \{2, 5, 10, 15, 20\}$. Only the selection $n_l = 2$ is inappropriate, as this value is too small to properly approximate the noise level.

4 | Data Preprocessing to Enable AFS Computations

Reliable estimation of the control parameters is necessary for computing meaningful AFS approximations. However, additional preprocessing of experimental data may be required before the AFS computation can begin. A common difficulty is baseline distortion. Even if the measured data D are nonnegative, baseline artifacts may prevent nonnegative factorizations and thus prevent the AFS computation. Or, as long as there is at least one nonnegative factorization, a situation with a hidden baseline artifact can easily remain undetected, and only a detailed manual analysis can reveal that no chemically realistic profiles may exist in a factorization. Another situation occurs when moderately relaxed nonnegativity constraints are satisfied in most frequency regions, but violated locally. Usual MCR factorizations rely on uniformly applying soft or hard constraints to ensure nonnegativity. Thus, the relaxation parameters respectively the setting of negative entries to zero are applied uniformly across all frequencies. Local baseline errors cannot be compensated for selectively without using weighted least squares.

In this section, we introduce an effective approach for overcoming the problem of hidden baseline artifacts that become amplified in factor profiles. Based on our experience with experimental data, we recommend applying this procedure before running an ambiguity analysis, if appropriate, no matter what method is used.

4.1 | Two-Stage Baseline Correction

In the first stage, a standard baseline correction is applied, for example by the rubber band method [42]. In the second stage, an initial pure component decomposition is computed. The resulting profiles may be chemically meaningful but still exhibit baseline distortions in one or more spectra. Let D denote the data after the first baseline correction and assume an initial factorization $D = CS^{\mathbf{T}}$. In the third step, a separate baseline correction is applied to $S^{\mathbf{T}}$. Broader regions with systematic negative offsets are shifted upward, and regions with systematic positive offsets are shifted downward. Let $\tilde{S}$ denote the corrected spectra and define $\Delta_S = S - \tilde{S}$. In the final step, the baseline contribution is removed from the data

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

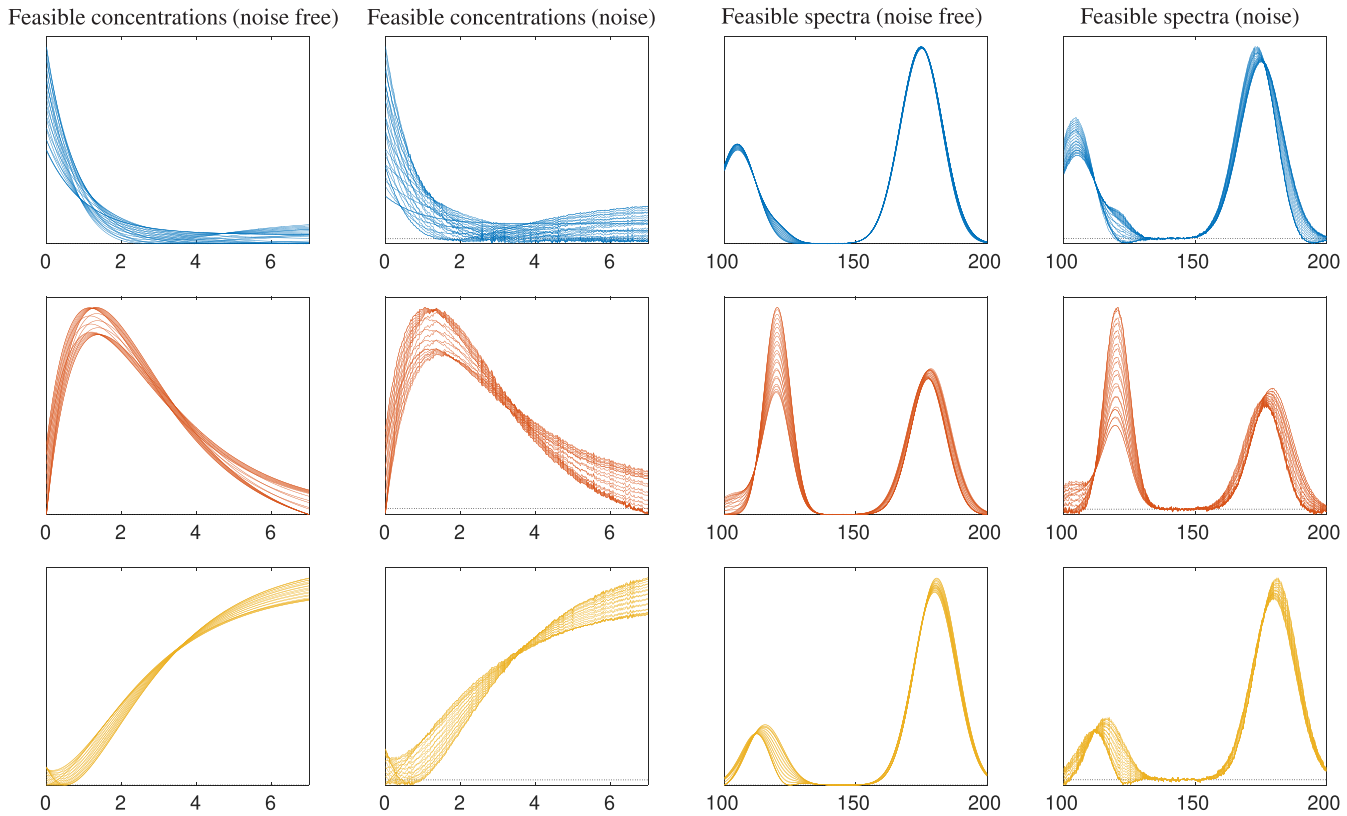


FIGURE 8 | The feasible bands for the model dataset 3. The bands for the noise-free data are plotted in Columns 1 and 3 and the bands for the data with noise are shown in Columns 2 and 4. These results coincide very well, demonstrating that the theoretically derived procedure produces reasonable approximations.

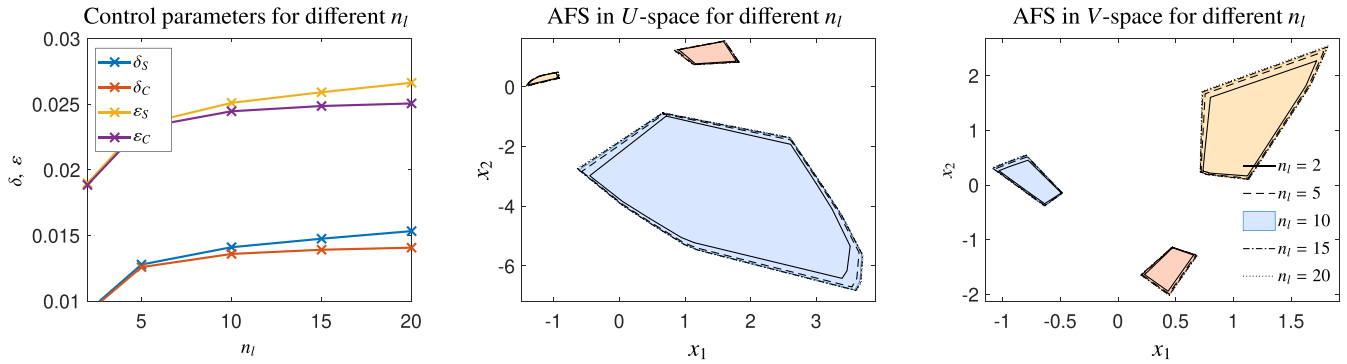


FIGURE 9 | The resulting control parameters and AFS sets for a variation of n_l. The parameter n_l defines the window size $2n_l + 1$ for the initial smoothing of the data values, by fitting a third-degree polynomial in a least squares sense. Left: Approximations of the control parameters and unbiased sample standard deviation in the time and frequency directions. Center and right: Approximations to the AFS sets using different values for n_l. The colored approximations correspond to $n_l = 10$ as in the lower plots of Figure 7. For comparison: the standard deviation for the noise of the model problem is 0.015.

via $\tilde{D} = D - C\Delta_S$. The workflow is illustrated schematically in Figure 10.

This procedure preserves information contained in higher singular values and related singular vectors, which would be lost if $C\tilde{S}^{\mathbf{T}}$ were formed directly. The rank of the data matrix is preserved. Moreover, the correction is not fully tied to the initial decomposition, so residual inconsistencies remain visible instead of being covered by the reconstructed product.

It is worth remark that this two-stage baseline correction is sometimes necessary for an ambiguity analysis, but it requires manual intervention.

Remark 2. If the AFS computation for a certain set of control parameters or the previous two-stage baseline correction is not successful, then a deeper analysis of the reasons is necessary. Possible reasons include non-random, systematic noise that cannot be removed by an additional baseline correction, as well as

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

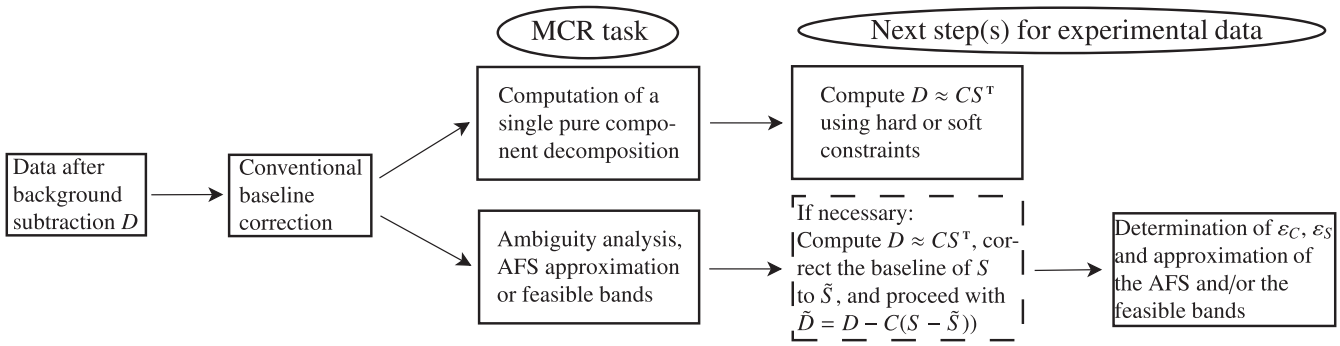


FIGURE 10 | Workflow to enable AFS computations for experimental data and the difference when computing only one MCR solutions. A conventional baseline correction enables the computation of initial profiles. A second baseline fine polishing removes systematic negative baseline artifacts and allows reasonable ambiguity analysis also for experimental data.

heteroscedastic or non-iid noise. Changing the initial factorization and decreasing the noise level can help to determine an improved AFS approximation in these cases.

4.2 | Application to Dataset 2

In Section 3.3, dataset 2 was used to illustrate the estimation of control parameters. However, a two-stage baseline correction was required to obtain meaningful AFS approximations. To clarify this point, we return to the data after the first baseline correction and present the intermediate results.

Figure 11 shows the data after background subtraction and after an initial baseline correction based on splines, similar to the rubber band method, see also Figure 2. In the original data, the baseline is located around -0.02, leading to more significant negative entries. After the first correction, the data are essentially nonnegative. Nevertheless, the ambiguity analysis remains unsatisfactory.

Applying the parameter estimation described above results in $\varepsilon_S = 1.89 \cdot 10^{-2}$ and $\varepsilon_C = 1.79 \cdot 10^{-2}$. These values appear moderate. However, the resulting AFS does not contain chemically meaningful profiles. Figure 12 shows the relaxed AFS sets together with a chemically interpretable factorization.

The difficulty is the following: There exist nonnegative factorizations that lack chemical interpretability. Conversely, chemically plausible factorizations exhibit broad negative regions in two pure component profiles. These negative contributions cancel in the product D and are therefore not visible in the data, but they appear in the factors. Increasing the relaxation parameters further enlarges the feasible region uniformly over all frequencies. The resulting AFS becomes excessively wide and loses interpretative value.

For the final preprocessing step, we use the computed factorization with chemically interpretable concentration profiles (center plots in Figure 12) and apply baseline fine polishing to the three associated pure component spectra. The corrected spectra, their deviations from the initial profiles, and the resulting data are shown in Figure 13.

The spectrum of the hydrido complex (yellow) is only slightly modified. The spectra of the acyl complex (blue) and the olefin

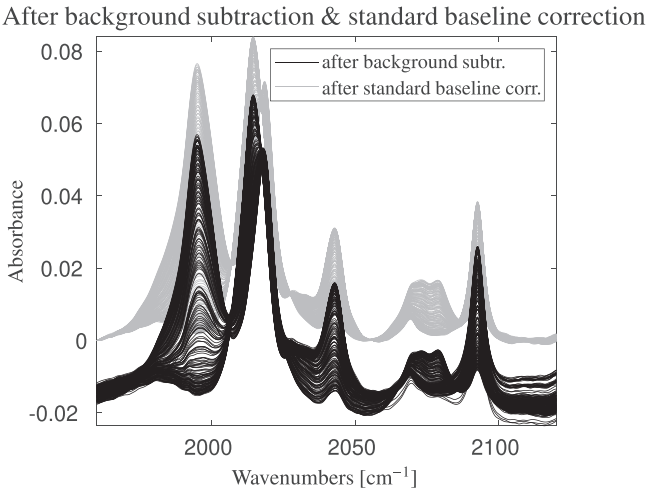


FIGURE 11 | The rhodium-catalyzed hydroformylation data after background subtraction. After applying a baseline correction similar to the rubber band method, but based on cubic spline functions, the data (gray lines and left plot in Figure 2) seems to be acceptable for an analysis. However, it does not allow for a suitable AFS computation because it is not nonnegatively factorizable and requires a baseline refinement, as shown in the right plot of Figure 13.

(red) are shifted upward in spectral regions with broad negative offsets. The spline-based rubber band method is applied again. The induced change in the data is small but essential for obtaining meaningful AFS approximations. The reference profiles are already shown in Figure 2.

Figure 14 shows the AFS sets of the corrected data together with the noise indicators ν_{ij}. An additional plot compares the singular values for three stages: no baseline correction, standard baseline correction, and subsequent baseline fine polishing. The final adjustment has no visible effect on the singular values. The data is only minimally modified and the data structure is preserved. In contrast, replacing the data by $D^{(\mathrm{new})} = C\tilde{S}^{\mathbf{T}}$, with $\tilde{S}$ as in Section 4.1, would produce a rank-3 matrix and remove higher-rank contributions unnecessarily.

The resulting control parameters are $\varepsilon_S = 1.84 \cdot 10^{-2}$ and $\varepsilon_C = 2.52 \cdot 10^{-2}$. As before, the values obtained from (8) are too small. The AFS set become meaningful after the parameters are increased by a factor of 4.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

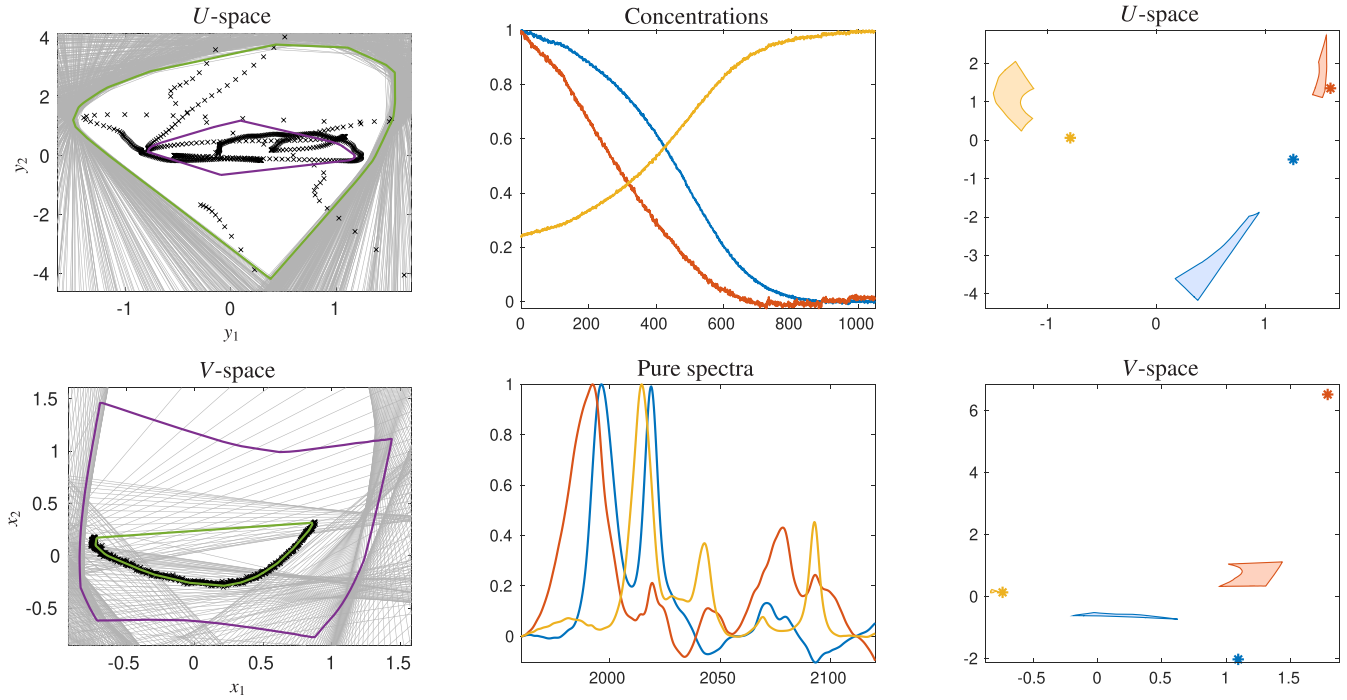


FIGURE 12 | The problem of the poor baseline of the IR dataset 2. After the initial baseline correction, the data does not contain significant negative entries (left plot in Figure 2). The inner and outer polytopes allow an enclosing triangle (left plots). However, the chemically meaningful factorization contains significant negative entries (center plots) and the AFS sets do not contain the true solution (right plots). The points representing the factorization from the middle plots are marked, but they are clearly outside of $\mathcal{M}_C$ and $\mathcal{M}_S$.

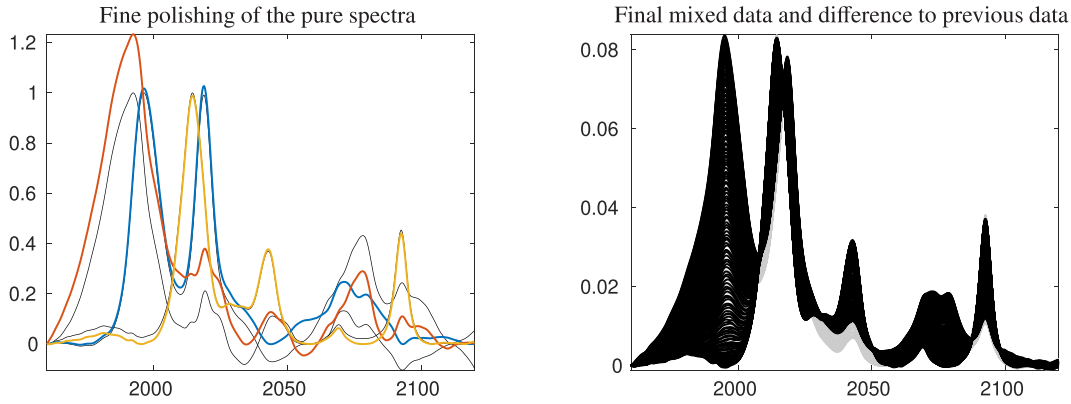


FIGURE 13 | The baseline refinement for the profiles in S of the factorization from Figure 12. Left: The new profiles are plotted in color, whereas the profiles from Figure 12 are plotted in black. Right: Data after the second baseline correction (black) and after the first correction (gray).

4.3 | Application to Dataset 3—Hidden Baseline Artifact

Next, we simulate a similar situation for the model dataset 3, to compare the results with the initial modification. We introduce an additional baseline distortion for the spectrum of species 2. As before, we add normally distributed random numbers with mean 0 and standard deviation 0.015 to D. The maximum baseline shift for $S(:,2)$ is -0.076, but the effect on D is minimal because the minimum entry before adding noise is -0.011. Thus, due to the additional noise, it is a hidden baseline problem. Figure 15 shows the new profiles. Similar to Section 3.4, we compute the AFS sets for $\varepsilon_C = \varepsilon_S = 0.025$ and obtain a factorization with negative entries no larger than required, see the upper plots of Figure 16. The problem with this factorization is that the second concentration profile is not meaningful because it is not unimodal with respect to noise. However, an alternative factorization with a unimodal profile for $C(:,1)$ exists and reveals the distortion of its corresponding spectral profile, see the lower plots in Figure 16. A final baseline correction of profile $S(:,2)$ does not result in a perfect fit of the initial displacement, as shown in Figure 17, but the approximation is good with respect to the random noise that is added to the mixed data, and it enables an ambiguity analysis with meaningful results.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

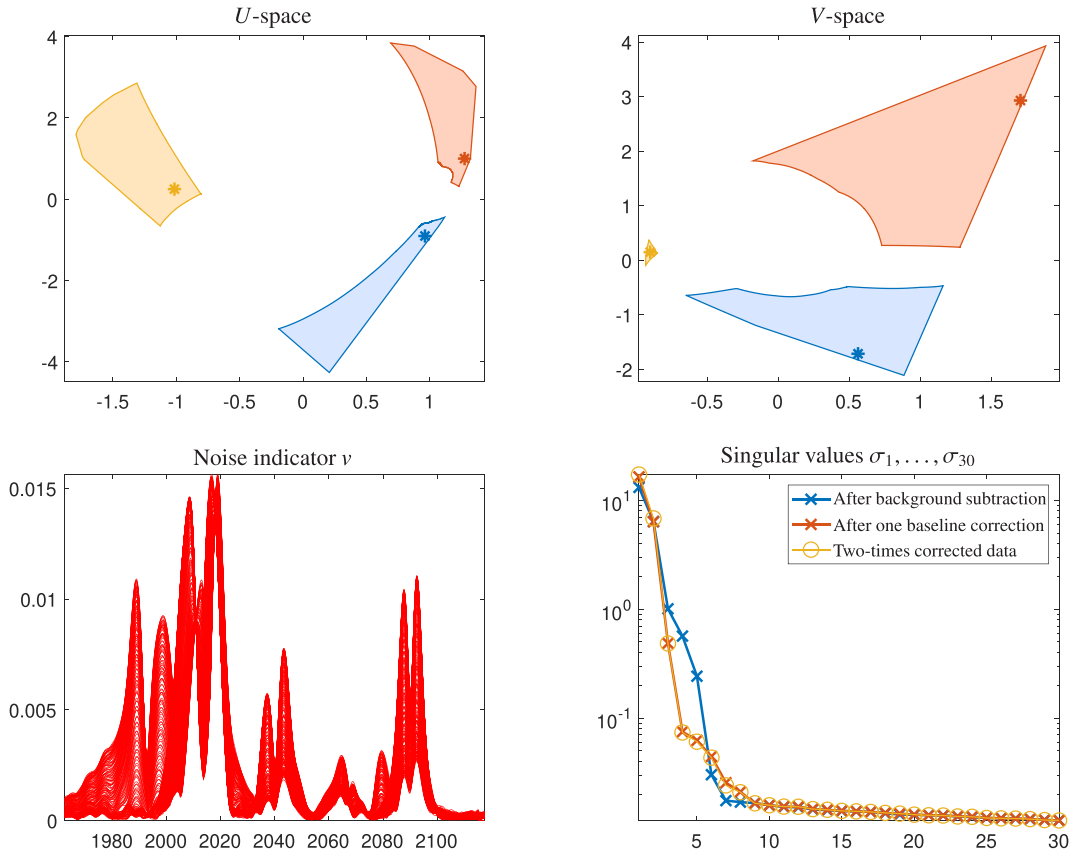


FIGURE 14 | Results after the two-stage baseline correction for dataset 2. Top row: The AFS sets, including the low-dimensional representations of the computed profiles (stars). Second row: Values for the unbiased sample variance v_{ij} from (6) in the frequency direction, as well as the singular values of the data matrix after background subtraction, after the first baseline correction, and after the second baseline correction.

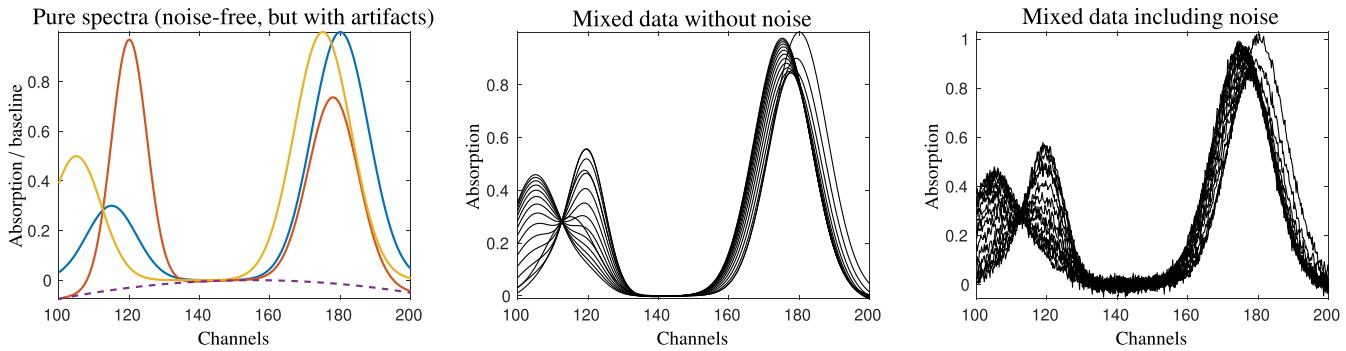


FIGURE 15 | Modified model problem from Section 4.3. Left: The pure spectra in blue, red (modified), and yellow, along with the new, artificially added baseline for $S(:,2)$ as a purple broken line. Center: the resulting mixed spectra, which are more or less nonnegative. Right: The noisy data.

5 | The Impact of Noise on Duality Relations

This section analyzes how perturbations in one factor influence the duality restrictions for the other factor. The columns of C and S, and their low-dimensional representations in the U- and V-spaces, are coupled by duality [16, 35–37]. The SVD and the matrix T in (1) are key to understanding duality because T and its invere are the coupling elements between C and S. A single profile in one factor restricts the admissible profiles in the other. In low-dimensional coordinates, namely, with respect to the bases of singular vectors, duality is expressed as follows. Let x according to (3) denote the representation of $S(:,i)$ in V-space and y the representation of $C(:,\ell)$ in U-space with $\ell \neq i$. Then

$$x^{\mathbf{T}}y = -1.$$

Hence, for a fixed profile $S(:,i)$ with representation $x \in \mathscr{M}$, all representations of $C(:,\ell)$, $\ell \neq i$, lie on the affine hyperplane

$$\mathscr{H}_x = \{y \in \mathbb{R}^{s-1} : x^{\mathbf{T}}y = -1\}$$

in U-space, see [16, 35, 37].

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

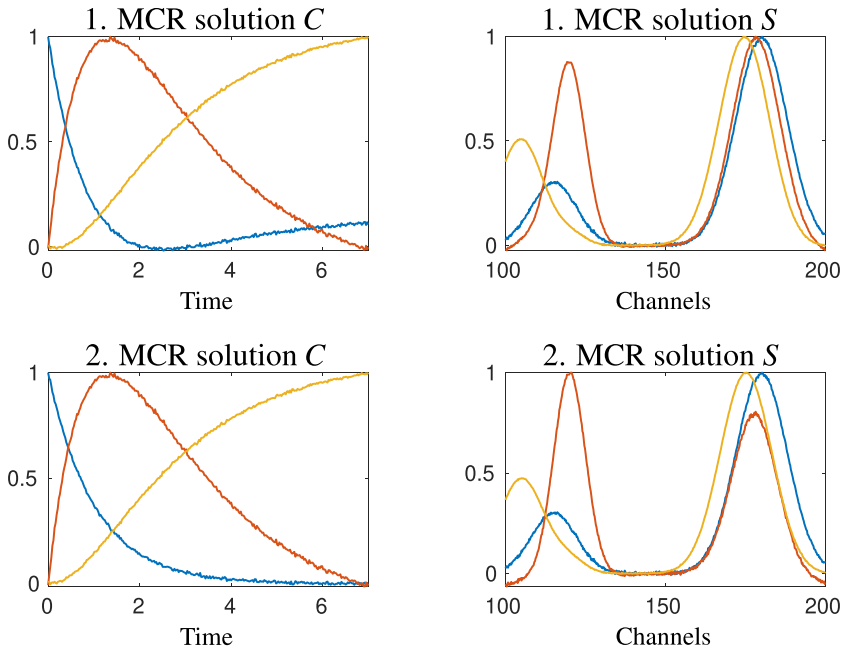


FIGURE 16 | Two factorizations of the distorted model dataset from Section 4.3. Upper: A feasible factorization but concentration profile 1 is not chemically meaningful, since not unimodal. Lower plots: a factorization with meaningful profiles for C but significant negative entries for the second pure spectrum—just representing the baseline distortion.

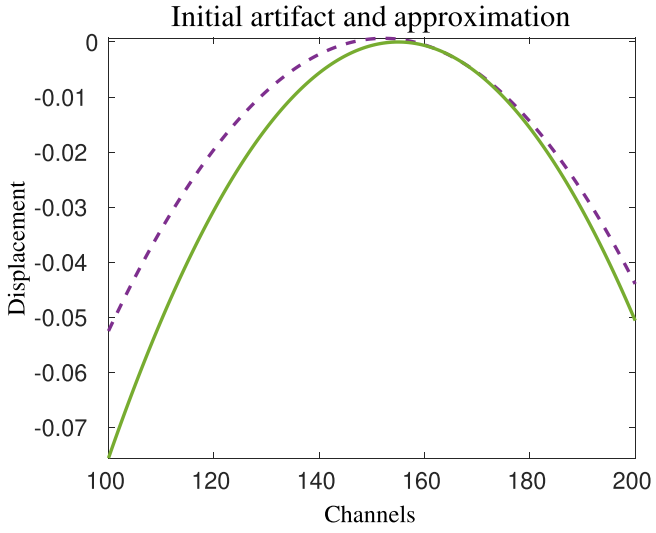


FIGURE 17 | The difference between the originally introduced baseline artifact for the red profile of Figure 15 (purple broken line) and its approximation (green line) using a two-stage baseline correction of $S(:,2)$ from the lower right plot of Figure 16. For the given perturbed data, the result appears to be acceptable.

We now assume that x is subject to uncertainty and study the resulting variation of $\mathscr{H}_x$. A complete error analysis cannot be presented here, even under assuming normal distributed noise; see the discussion in Section 5.4. Instead, the effect of perturbations is examined by a case study based on random noise.

5.1 | Noise Models

To model uncertainty, assume that a pure component spectrum a, for example obtained from a separate measurement, is affected by noise. Let

$$\tilde{a} = a + e$$

denote the perturbed spectrum, where $e \in \mathbb{R}^n$ represents the noise.

The first question concerns the choice of a suitable noise model. One option is to assume bounded perturbations and model each component by a β-distribution

$$e_j \sim \text{Beta}(p, q).$$

This corresponds to prescribing an uncertainty band around a, so that $\tilde{a}$ remains within fixed bounds. Such a model represents measurement uncertainty by a tube around a and allows one to study its effect on the dual hyperplanes in the U-space; see the left plot in Figure 18.

Alternatively, the perturbation can be modeled by normally distributed noise with mean zero and variance σ^2,

$$e \in \mathbb{R}^n, \qquad e_j \sim \mathcal{N}(0, \sigma^2).$$

This model reflects random measurement noise without explicit bounds.

5.2 | A Case Study Using Dataset 1

The case study is based on dataset 1. The procedure is as follows. Each pure component profile $S(:,i)$ is perturbed separately. For every profile, $N = 500$ noisy realizations $\tilde{a}$ are generated using random perturbations. For each realization, we compute

1. the low-dimensional representation in V-space,
2. the corresponding dual hyperplane (line) in U-space.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The figures referenced below only display the outer envelopes. They illustrate the range induced by the perturbations in both spaces. This approach simulates an uncertainty tube around the reference profile, in particular when using the β-distribution, and visualizes its effect on the dual constraints.

5.3 | Results Using the β-Distribution

The true factor $S^{\mathbf{T}}$ is taken as reference, and perturbations are applied directly to it using the β-distribution with mean 0 and bandwidth $2\epsilon/\max(S(:,i))$ for $\epsilon = 0.03$. For each profile $S(:,i)$, $i = 1, 2, 3$, the noisy realizations are generated as

$$\tilde{S}_{ji} = S_{ji} + \frac{\epsilon}{\|S(:,i)\|_\infty} e_j, \qquad e_j \sim \text{Beta}(0.1, 0.1),$$

repeated N times.

The perturbed spectra, their low-dimensional representations, and the corresponding envelopes in the dual space are shown in Figure 18. The perturbation produces a band of width 6 % of the maximal absorption around each profile. The low-dimensional representations in V-space form approximately circular regions. The envelopes of the associated dual hyperplanes in U-space exhibit curved boundaries resembling rotated and skewed hyperbolas. These regions may be interpreted as uncertainty sets (or sets of confidence) around the unperturbed representations, although not in a formal statistical sense.

5.4 | Results Using the Normal Distribution

Next, noise is modeled by a normal distribution with mean 0 and variance $\sigma^2 = 0.025$. Although this distribution is analytically convenient, it does not allow a direct characterization of the induced distribution in the low-dimensional coordinates.[1] Therefore, the perturbations are generated numerically and their effects in the V- and U-spaces are examined.

Figure 19 shows the results. They are similar to those obtained with the β-distribution. The low-dimensional representations in V-space form approximately ellipsoidal regions that are slightly stretched in radial direction from the origin.

5.5 | Different Sensitivities in the Dual Space

Figures 18 and 19 show that dual hyperplanes are particularly sensitive when the corresponding point in the V-space is close to

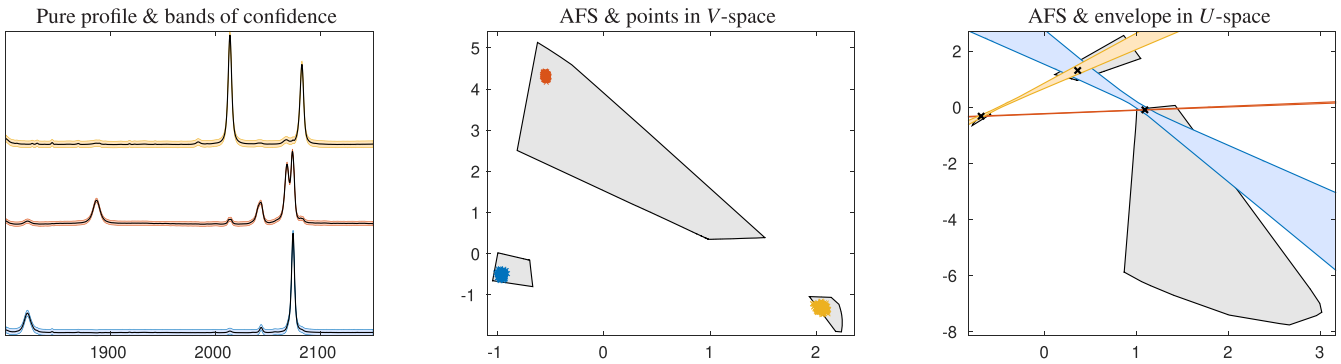


FIGURE 18 | Impact of noise in the dual space for dataset 1 using the β-distribution. Left: Perturbations generated with $p = q = 0.1$ and bandwidth $2\epsilon/\max(S(:,i))$ for the respective spectrum, simulating an uncertainty tube around the solution. Center: Corresponding low-dimensional representations in V-space. Right: Envelopes of the resulting dual hyperplanes in U-space. The effect is significantly smaller for the red spectrum than for the blue spectrum. This difference reflects the relative positions of their low-dimensional representations in V-space. In each experiment, only one spectrum is perturbed, whereas the remaining spectra are kept fixed.

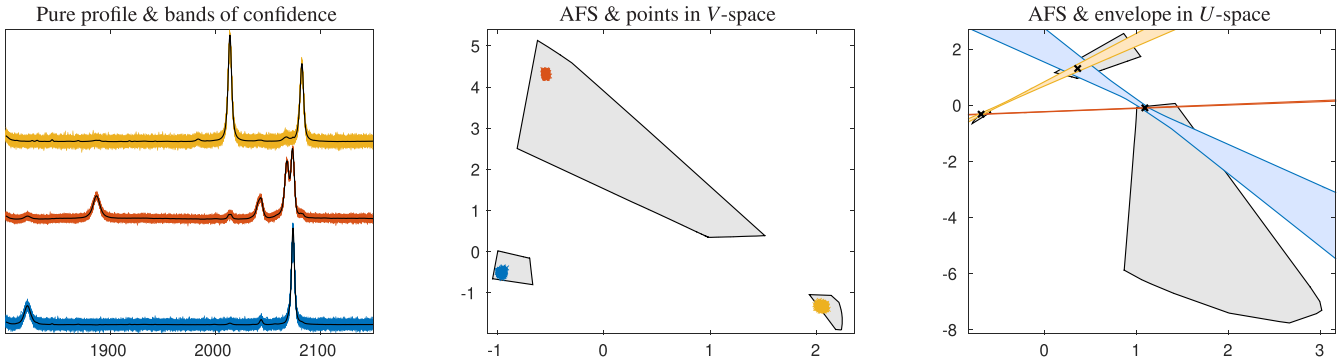


FIGURE 19 | The impact of noise in the dual space for dataset 1 using the standard normal distribution to model noise. Left: each pure component spectrum is perturbed $N = 500$ times. Center: Low-dimensional representations in the V-space. Right: The envelopes of the dual hyperplanes in the U-space corresponding to the perturbed points.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the origin. Consequently, duality restrictions derived from measured spectra with small $\|x\|$ are more affected by perturbations than those associated with points farther from the origin. This behavior is consistent with Lemma 3.7 in [16]: Let x be a point in V-space, y a point on the dual hyperplane $\mathscr{H}_x$ in U-space, and δ_x a perturbation of x. Neglecting quadratic terms, the perturbation δ_y in U-space satisfies

$$\|\delta_y\| \geq \frac{|\delta_x^{\mathrm{T}} y|}{\|x\|}.$$

Hence, the sensitivity increases as $\|x\|$ decreases.

A further relation concerns angles. Lemma 3.8 in [16] for $s = 3$ and Lemma 3.79 in [43] show that the angle between x and $x + \delta_x$ in V-space equals the angle between the corresponding dual hyperplanes $\mathscr{H}_x$ and $\mathscr{H}_{x+\delta_x}$ in U-space,

$$\measuredangle\left(x, x+\delta_x\right) = \measuredangle\left(\mathscr{H}_x, \mathscr{H}_{x+\delta_x}\right).$$

Both effects are illustrated in Figure 20. Note that in both U- and V-space the y_2 and x_2 axes, respectively, have a larger scaling than the y_1 and x_1 axes, which influences the visual impression of sensitivity.

6 | Sensitivity Indicator for the Data Points

This section analyzes the sensitivity of individual data points of the mixed spectra with respect to noise. The distributions of the data points in the U- and V-spaces in Figures 4 and 6 indicate substantial differences in noise sensitivity. These points are the low-dimensional representations of the columns and rows of D. If a row or column of D contains entries close to zero, its representation is typically more sensitive to perturbations. This phenomenon is known from the concept of essential data points [44, 45]. For noisy data, essential points are more reliably identified by numerical criteria than by direct geometric conditions [46].

The data points in the V-space are given by a_i in (4), and the data points in U-space are b_i with

$$b_i = \frac{(V\Sigma^{1/2})^{\mathrm{T}}(2{:}s, i)}{(V\Sigma^{1/2})^{\mathrm{T}}(1, i)}, \quad i = 1, \ldots, n.$$

For the Euclidean norm of a_i, and thus its distance from the origin, we have

$$\|a_i\|_2 = \frac{\left\| U(i, 2{:}s) \cdot \mathrm{diag}\left(\sqrt{\sigma_2}, \ldots, \sqrt{\sigma_s}\right)\right\|_2}{U_{i1}\sqrt{\sigma_1}}$$

with the singular values $\sigma_1, \ldots, \sigma_s$ of D. (The first right singular vector $U(:, 1)$ can be assumed component-wise strictly positive for an irreducible matrix $D^{\mathrm{T}}D$.) If, for a given i, the denominator $U_{i1}\sqrt{\sigma_1}$ is small in magnitude while at least one numerator $U_{ij}\sqrt{\sigma_j}$, $j \in 2, \ldots, s$, is not small, then $\|a_i\|_2$ becomes large. In this case, the point a_i is likely to lie far from the origin and may violate the constraint $a_i \in \mathscr{F}_S$.

To quantify this effect, we introduce the sensitivity indicator $e^{(a)} \in \mathbb{R}^k$ defined by

$$e_i^{(a)} = \max_{j=2,\ldots,s} \left| \frac{U_{ij}\sqrt{\sigma_j}}{U_{i1}\sqrt{\sigma_1}} \right|, \quad i = 1, \ldots, k. \tag{9}$$

Large values of $e_i^{(a)}$ indicate increased sensitivity of the corresponding data point in V-space.

Analogously,

$$\|b_i\|_2 = \frac{\left\| V(i, :) \cdot \mathrm{diag}\left(\sqrt{\sigma_2}, \ldots, \sqrt{\sigma_s}\right)\right\|_2}{V_{i1}\sqrt{\sigma_1}}$$

and we define the sensitivity indicator $e^{(b)} \in \mathbb{R}^n$ by

$$e_i^{(b)} = \max_{j=2,\ldots,s} \left| \frac{V_{ij}\sqrt{\sigma_j}}{V_{i1}\sqrt{\sigma_1}} \right|, \qquad i = 1, \ldots, n, \tag{10}$$

As before, large values of $e_i^{(b)}$ indicate higher sensitivity of the corresponding data point in the U-space to noise.

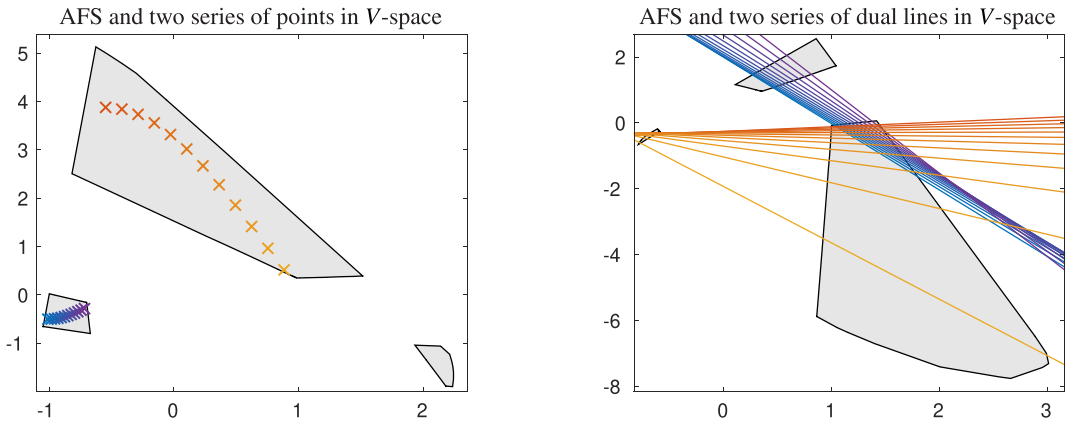


FIGURE 20 | The sensitivity of the duality restriction in the U-space depends on the distance of the points from the origin in the V-space, and vice versa. Left: two series of points in V-space for dataset 1. Right: the associated dual lines in U-space. Small changes to points far away from the origin result in small changes of the dual lines, whereas small changes to points close to the origin result in larger changes to the dual lines.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

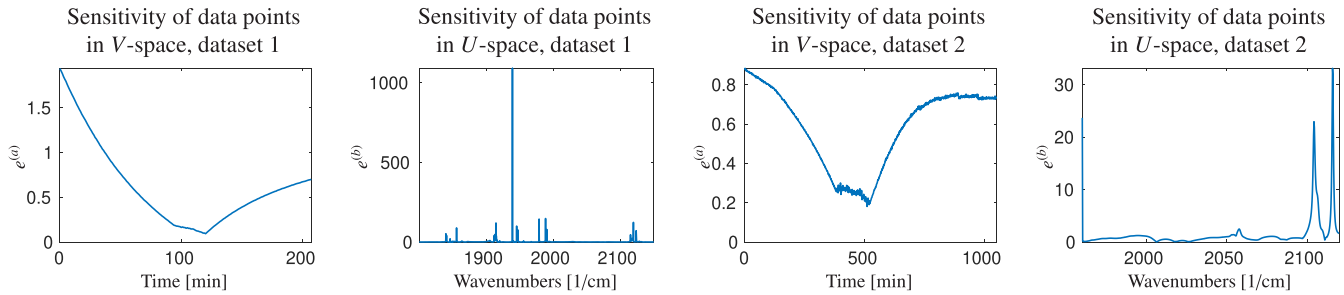


FIGURE 21 | Sensitivity indicators for the data points in the datasets 1 and 2. Small values in $e^{(a)}$ indicate a low sensitivity of the data points in V-space, whereas some high values in $e^{(b)}$ indicate a high sensitivity of the data points in U-space. See Figures 4 and 6 for more information.

In IR or Raman data, this situation occurs when $V_{i1}\sqrt{\sigma_1}$ is small while at least one term $V_{ij}\sqrt{\sigma_j}$, $2 \le j \le s$, remains significant. Figure 21 displays the sensitivity indicators for datasets 1 and 2. For both datasets, the indicators $e^{(a)}$ are small, showing that the data points a_i in V-space are relatively stable. In contrast, several entries of $e^{(b)}$ are considerably larger, indicating that the data points in U-space are more sensitive to perturbations.

The indicators $e^{(a)}$ and $e^{(b)}$ characterize the sensitivity of the data points and thus of the inner polytope. They also reflect the sensitivity of the outer polytope in the dual space because data points are dual to facets of the dual outer polytope. For the outer polytopes, the roles of V- and U-space are interchanged; see Figures 4 and 6. For both datasets, the facet-defining conditions of the outer polytopes (gray lines) are much more sensitive in V-space than in U-space. One potential use of the indicators is to identify outliers and distinguish them from absorptions of minor components, see e.g., [47]. If necessary, the identified spectra or frequencies can be removed from the data.

7 | Outlook

The selection of suitable control parameters is decisive for analyzing the rotational ambiguity of an MCR solution. This paper proposes procedures to approximate the AFS based on statistical arguments and on our practical experience from applying AFS computations to experimental data over the past years. The proposed rules provide guidance but do not constitute a universal method. Parameter values must be chosen and adapted to the specific dataset. It is advisable to vary them within a reasonable range and assess the stability of the resulting AFS. The presented examples show that the approach yields meaningful approximations in typical situations.

The procedure does not compensate for an inadequate baseline correction. Baseline distortions are systematic effects, whereas the aim here is to quantify and accommodate predominantly random noise. Further aspects remain open. In systems with large differences in maximal absorption between components, minor species are often more strongly affected by noise. Their profiles are typically associated with later singular vectors and are therefore more sensitive to perturbations than dominant components. This issue will be addressed in a forthcoming paper. The influence of systematic, non-random noise is also not considered here. If such effects are pronounced, and if the noise levels are too high, it may be impossible to make a reasonable approximation of the AFS sets and feasible bands.

Acknowledgements

Open Access funding enabled and organized by Projekt DEAL.

Data Availability Statement

The authors choose not to share the data.

Endnotes

[1] The difficulty arises from numerical integration when computing the expectation of the low-dimensional representation. Scaling by 1 in the first column of T and dividing by the contribution of the first singular vector may result in division by 0.

References

1. M. Maeder and Y. M. Neuhold. *Practical Data Analysis in Chemistry, Volume 26 of Data Handling in Science and Technology* (Elsevier, 2007).

2. E. R. Malinowski, *Factor Analysis in Chemistry* (Wiley, 2002).

3. H. Abdollahi and R. Tauler, "Uniqueness and Rotation Ambiguities in Multivariate Curve Resolution Methods," *Chemometrics and Intelligent Laboratory Systems* 108, no. 2 (2011): 100–111.

4. M. Sawall, H. Schröder, D. Meinhardt, and K. Neymeyr, "On the Ambiguity Underlying Multivariate Curve Resolution Methods," in *Comprehensive Chemometrics: Chemical and Biochemical Data Analysis*, eds. R. Tauler, B. Walczak, and S. Brown (Elsevier, 2020), 199–231.

5. P. J. Gemperline, "Computation of the Range of Feasible Solutions in Self-Modeling Curve Resolution Algorithms," *Analytical Chemistry* 71, no. 23 (1999): 5398–5404.

6. J. Jaumot and R. Tauler, "MCR-BANDS: A User Friendly MATLAB Program for the Evaluation of Rotation Ambiguities in Multivariate Curve Resolution," *Chemometrics and Intelligent Laboratory Systems* 103, no. 2 (2010): 96–107.

7. R. Tauler, "Calculation of Maximum and Minimum Band Boundaries of Feasible Solutions for Species Profiles Obtained by Multivariate Curve Resolution," *Journal of Chemometrics* 15, no. 8 (2001): 627–646.

8. A. C. Olivieri, "Estimating the Boundaries of the Feasible Profiles in the Bilinear Decomposition of Multi-Component Data Matrices," *Chemometrics and Intelligent Laboratory Systems* 216 (2021): 104387.

9. A. C. Olivieri, M. Sawall, K. Neymeyr, and R. Tauler, "Noise Effects on Band Boundaries in Multivariate Curve Resolution of Three-Component

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Systems," *Chemometrics and Intelligent Laboratory Systems* 228 (2022): 104636.

10. O. S. Borgen and B. R. Kowalski, "An Extension of the Multivariate Component-Resolution Method to Three Components," *Analytica Chimica Acta* 174 (1985): 1–26.

11. R. Rajkó and K. István, "Analytical Solution for Determining Feasible Regions of Self-Modeling Curve Resolution (SMCR) Method Based on Computational Geometry," *Journal of Chemometrics* 19, no. 8 (2005): 448–463.

12. A. Jürß, M. Sawall, and K. Neymeyr, "On Generalized Borgen Plots. I: From Convex to Affine Combinations and Applications to Spectral Data," *Journal of Chemometrics* 29, no. 7 (2015): 420–433.

13. A. Golshan, H. Abdollahi, and M. Maeder, "Resolution of Rotational Ambiguity for Three-Component Systems," *Analytical Chemistry* 83, no. 3 (2011): 836–841.

14. A. Golshan, M. Maeder, and H. Abdollahi, "Determination and Visualization of Rotational Ambiguity in Four-Component Systems," *Analytica Chimica Acta* 796 (2013): 20–26.

15. M. Sawall, C. Kubis, D. Selent, A. Börner, and K. Neymeyr, "A Fast Polygon Inflation Algorithm to Compute the Area of Feasible Solutions for Three-Component Systems. I: Concepts and Applications," *Journal of Chemometrics* 27, no. 5 (2013): 106–116.

16. M. Sawall and K. Neymeyr, "On the Area of Feasible Solutions and Its Reduction by the Complementarity Theorem," *Analytica Chimica Acta* 828 (2014): 17–26.

17. M. Sawall and K. Neymeyr, "A Ray Casting Method for the Computation of the Area of Feasible Solutions for Multicomponent Systems: Theory, Applications and FACPACK-Implementation," *Analytica Chimica Acta* 960 (2017): 40–52.

18. R. Laursen and A. Hobolth, "A Sampling Algorithm to Compute the Set of Feasible Solutions for Nonnegative Matrix Factorization With an Arbitrary Rank," *SIAM Journal on Matrix Analysis and Applications* 43, no. 1 (2022): 257–273.

19. M. Sawall, T. Andersons, C. Wei, C. Kubis, and K. Neymeyr, "Sampling-Based Computation of the Sets of Feasible Solutions and Feasible Bands for Noisy Data," *Chemometrics and Intelligent Laboratory Systems* 268 (2025): 105565.

20. A. Golshan, H. Abdollahi, S. Beyramysoltan, et al., "A Review of Recent Methods for the Determination of Ranges of Feasible Solutions Resulting From Soft Modelling Analyses of Multivariate Data," *Analytica Chimica Acta* 911 (2016): 1–13.

21. M. Sawall, A. Moog, and K. Neymeyr. "FACPACK: A Software for the Computation of Multi-Component Factorizations and the Area of Feasible Solutions," Revision 1.3. FACPACK homepage: http://www.math.uni-rostock.de/facpack/, 2018.

22. W. Chew, E. Widjaja, and M. Garland, "Band-Target Entropy Minimization (BTEM): An Advanced Method for Recovering Unknown Pure Component Spectra. Application to the FT-IR Spectra of Unstable Organometallic Mixtures," *Organometallics* 21, no. 9 (2002): 1982–1990.

23. K. Neymeyr, M. Sawall, and D. Hess, "Pure Component Spectral Recovery and Constrained Matrix Factorizations: Concepts and Applications," *Journal of Chemometrics* 24, no. 2 (2010): 67–74.

24. E. Widjaja, C. Li, W. Chew, and M. Garland, "Band Target Entropy Minimization. A Robust Algorithm for Pure Component Spectral Recovery. Application to Complex Randomized Mixtures of Six Components," *Analytical Chemistry* 75, no. 17 (2003): 4499–4507.

25. B. N. Leidecker, D. P. Fuentes, C. Wei, et al., "In Situ FTIR Spectroscopic Investigations on Rhodium Carbonyl Complexes in the Absence of Phosphorus Ligands Under Hydroformylation Conditions," *New Journal of Chemistry* 48 (2024): 18365–18375.

26. C. Kubis, M. Sawall, A. Block, et al., "An Operando FTIR Spectroscopic and Kinetic Study of Carbon Monoxide Pressure Influence on Rhodium-Catalyzed Olefin Hydroformylation," *Chemistry—A European Journal* 20, no. 37 (2014): 11921–11931.

27. C. Kubis, D. Selent, M. Sawall, et al., "Exploring Between the Extremes: Conversion Dependent Kinetics of Phosphite-Modified Hydroformylation Catalysis," *Chemistry—A European Journal* 18, no. 28 (2012): 8780–8794.

28. J. Jaumot, A. de Juan, and R. Tauler, "MCR-ALS GUI 2.0: New Features and Applications," *Chemometrics and Intelligent Laboratory Systems* 140 (2015): 1–12.

29. H. Kim and H. Park, "Nonnegative Matrix Factorization Based on Alternating Nonnegativity Constrained Least Squares and Active Set Method," *SIAM Journal on Matrix Analysis and Applications* 30, no. 2 (2008): 713–730.

30. D. D. Lee and H. S. Seung, "Learning the Parts of Objects by Non-Negative Matrix Factorization," *Nature* 401, no. 6755 (1999): 788–791.

31. M. Sawall and K. Neymeyr, "A Fast Polygon Inflation Algorithm to Compute the Area of Feasible Solutions for Three-Component Systems. II: Theoretical Foundation, Inverse Polygon Inflation, and FACPACK Implementation," *Journal of Chemometrics* 28, no. 5 (2014): 633–644.

32. P. J. Gemperline and E. Cash, "Advantages of Soft Versus Hard Constraints in Self-Modeling Curve Resolution Problems. Alternating Least Squares With Penalty Functions," *Analytical Chemistry* 75, no. 16 (2003): 4236–4243.

33. G. W. Stewart, "Determining Rank in the Presence of Error," Technical report, University of Maryland, College Park, MD, USA, 1992.

34. G. W. Stewart, "On the Early History of the Singular Value Decomposition," *SIAM Review* 35 (1993): 551–556.

35. R. C. Henry, "Duality in Multivariate Receptor Models," *Chemometrics and Intelligent Laboratory Systems* 77, no. 1–2 (2005): 59–63.

36. R. Rajkó, "Natural Duality in Minimal Constrained Self Modeling Curve Resolution," *Journal of Chemometrics* 20, no. 3–4 (2006): 164–169.

37. M. Sawall, A. Jürß, H. Schröder, and K. Neymeyr, "Simultaneous Construction of Dual Borgen Plots. I: The Case of Noise-Free Data," *Journal of Chemometrics* 31, no. 12 (2017): e2954.

38. P. J. Gemperline and M. A. Lakeh, "Chapter 1—Soft Constraints in Curve Resolution Problems," in *Machine Learning and Pattern Recognition Methods in Chemistry From Multivariate and Data Driven Modeling*, ed. J. B. Ghasemi (Elsevier, 2023), 1–26.

39. A. C. Olivieri, K. Neymeyr, M. Sawall, and R. Tauler, "How Noise Affects the Band Boundaries in Multivariate Curve Resolution," *Chemometrics and Intelligent Laboratory Systems* 220 (2022): 104472.

40. I. Bronstein, K. Semendjajew, G. Musiol, and H. Mühlig, *Handbook of Mathematics* (Springer Berlin Heidelberg, 2015).

41. A. Savitzky and M. J. E. Golay, "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Analytical Chemistry* 36, no. 8 (1964): 1627–1639.

42. S. Wartewig, ed., *IR and Raman Spectroscopy: Fundamental Processing. Spectroscopic Techniques* (Wiley-VCH, 2003).

43. M. Sawall, "Analyse und Berechnung Niedrigdimensionaler Darstellungen von Lösungsmengen zur Nichtnegativen Matrixfaktorisierung," Habilitation thesis, Universität Rostock, 2019.

44. M. Ghaffari, N. Omidikia, and C. Ruckebusch, "Essential Spectral Pixels for Multivariate Curve Resolution of Chemical Images," *Analytical Chemistry* 91, no. 17 (2019): 10943–10948.

45. M. Sawall, C. Kubis, H. Schröder, et al., "Multivariate Curve Resolutions Methods and the Design of Experiments," *Journal of Chemometrics* 34, no. 2 (2020): e3159.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

46. M. Sawall, C. Ruckebusch, M. Beese, R. Francke, A. Prudlik, and K. Neymeyr, "An Active Constraint Approach to Identify Essential Spectral Information in Noisy Data," *Analytica Chimica Acta* 1233 (2022): 340448.

47. R. Vitale and C. Ruckebusch, "On a Black Hole Effect in Bilinear Curve Resolution Based on Least Squares," *Journal of Chemometrics* 37, no. 2 (2023): e3442.

1099128x, 2026, 10, Downloaded from https://analyticalsciencejournals.onlinelibrary.wiley.com/doi/10.1002/cem.70179 by UNR - Univ Nacional de Rosario, Wiley Online Library on [22/10/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License